Computational and Data Grids in Large-Scale Science and Engineering

William E. Johnston[a]
Lawrence Berkeley National Laboratory and NASA Ames Research Center

Keywords

Grids; heterogeneous, widely distributed computing; NASA's Information Power Grid (IPG); DOE Science Grid; Grid applications

Abstract

As the practice of science moves beyond the single investigator due to the complexity of the problems that now dominate science, large collaborative and multi-institutional teams are needed to address these problems.

In order to support this shift in science, the computing and data handling infrastructure that is essential to most of modern science must also change in order to support this increased complexity. This is the goal of computing and data Grids: Software infrastructure that facilitates solving large-scale problems by providing the mechanisms to access, aggregate, and manage the computer network based infrastructure of science. This infrastructure includes computing systems, data archive systems, scientific instruments, and computer mediated human collaborations.

This paper examines several large-scale science problems, their requirements for computing and data Grid infrastructure, and the current approaches to providing the necessary functionality.

1 Introduction

"Grids" (see [1]) are an approach for building dynamically constructed problem solving environments using geographically and organizationally dispersed high performance computing and data handling resources.

Functionally, Grids are tools, middleware, and services for

- providing a uniform look and feel to a wide variety of distributed computing and data resources
- supporting construction, management, and use of widely distributed application systems
- facilitating human collaboration and remote access to, and operation of, scientific and engineering instrumentation systems
- managing and securing this computing and data infrastructure as a persistent service

This is accomplished through a set of uniform software services (the Common Grid Services - described in more detail below) that manage and provide access to heterogeneous, distributed resources. These services may be summarized as:

- information services for resource discovery
- resource co-scheduling
- authentication and authorization
- auditing
- global event services
- data cataloguing, publishing, and subscribing
- collaboration and remote instrument management and access services
- communication services
- resource specification and request
- uniform data access
- security services
- monitoring
- global queuing
- resource brokering
- data location management
- fault management

The overall motivation for the current large-scale (multi-institutional) Grid projects is to enable the resource interactions that facilitate large-scale science and engineering such as aerospace systems design, high energy physics data analysis, climatology, large-scale remote instrument operation, etc.

One motivation for computing, data, and instrument Grids is that they will provide significant new capabilities to scientists and engineers by facilitating routine construction of information based problem solving environments that

[a] wejohnston@lbl.gov, www.itg.lbl.gov/~wej, wej@nas.nasa.gov, www.ipg.nasa.gov

are built on-demand from large pools of resources. That is, Grids will routinely – and easily, from the user's point of view – facilitate applications such as:

- coupled, multidisciplinary simulations too large for single computing systems (e.g., multi-component turbomachine simulation – see [2] and [3]
- management of very large parameter space studies where thousands of low fidelity simulations explore, e.g., the aerodynamics of the next generation space shuttle in its many operating regimes (from Mach 27 at entry into the atmosphere to landing)
- use of widely distributed, federated data archives (e.g., simultaneous access to metrological, topological, aircraft performance, and flight path scheduling databases supporting a National Air Transportation Simulation system)
- coupling large-scale computing and data systems to scientific and engineering instruments so that complex real-time data analysis results can be used by the experimentalist in ways that allow direct interaction with the experiment (e.g. Cosmology data analysis involving telescope and satellite interaction, and coupling to simulations)
- single computational problems too large for any single system (e.g. extremely high resolution rotocraft aerodynamic calculations)

2 Motivating Applications

As the problem tackled by the science and research engineering communities become more and more complex, the computing requirements are not just for more computing power, but for dealing with more complex application systems as well. We present two representative applications. Both require high capability computing and data handling, and also require a complex mix or resources – multiple computers, databases, archives, instruments, etc., all of which must be carefully coordinated to solve the problem.

2.1 NASA's Aviation Safety Program

A current NASA R&D[a] project is to develop the approach and technology for modeling the entire commercial airspace of the US; that is, to produce a virtual national air space. The benefits of this range from potentially much more efficient utilization of airports and flight paths, to determining, and possibly correcting, aircraft related emergency conditions while in flight. The modeling involves integrating huge amounts of flight and ground operations data, weather, terrain, etc., together with whole aircraft simulations of the approximately 22,000 commercial flights per day in the US.

Most of the aircraft sub-systems (engine, wing lift, control surfaces, landing gear, etc.) are well studied individually, but combining these into a whole system simulation of the aircraft, and then integrating the result into an operational air space, is a considerable challenge.

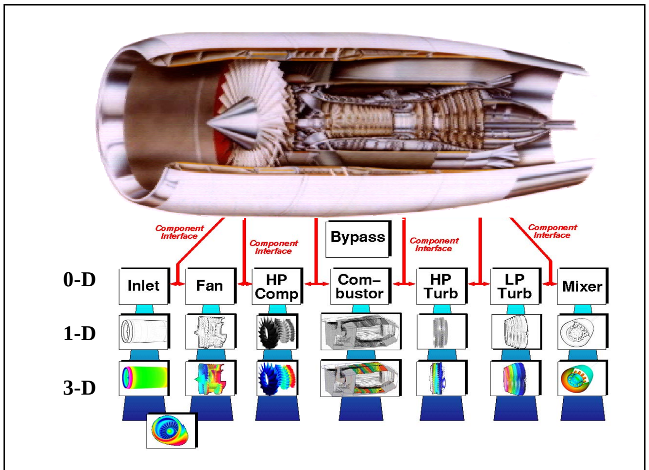

Figure 1. NPSS Simulation Architecture

NPSS supplies tools for integrating codes, accessing geometry, zooming, coupling, and security. (Figure courtesy Gregory J. Follen, Computing & Interdisciplinary Systems Office, NASA Glenn Research Center.)

The NPSS program [2] at NASA Glenn is working on coupling the many component models required to simulate an operational jet engine (see Figure 1), and integrating the resulting engine model with operational data. The sub-system simulations have been developed over a long time and they are written in a variety of languages (e.g. FORTRAN) and in a variety of styles. The NPSS program has built an application framework for coupling these together [3], and this approach is being extended to the whole aircraft problem.

[a] The "Virtual National Air Space" is the vision of Yuri Gawdiak and Bill McDermott, NASA Ames, and John Lytle and Gregory Follen, NASA Glenn.

Inserting the combined simulations into an operational environment means that the model drivers – initial conditions, boundary conditions, forcing functions, etc., – are now derived from various environmental data – observed velocity, elevation, pilot's throttle setting, etc. This requires changing from static input files to potentially dynamic databases or live data feeds, and perhaps having to build secondary models that convert observed quantities into the quantities needed to drive, e.g., the engine simulations.

Figure 2 illustrates the end-game in which all of the aircraft are simulated in the National air space, and these simulations are combined with the operational data to produce the Virtual National Air Space[a]. Such a whole systems simulation will clearly involve managing and coordinating a large number of organizationally and geographically dispersed computing and data resources: The computing power, data archives, special databases maintained by discipline experts, etc., will never be in one location or institution.

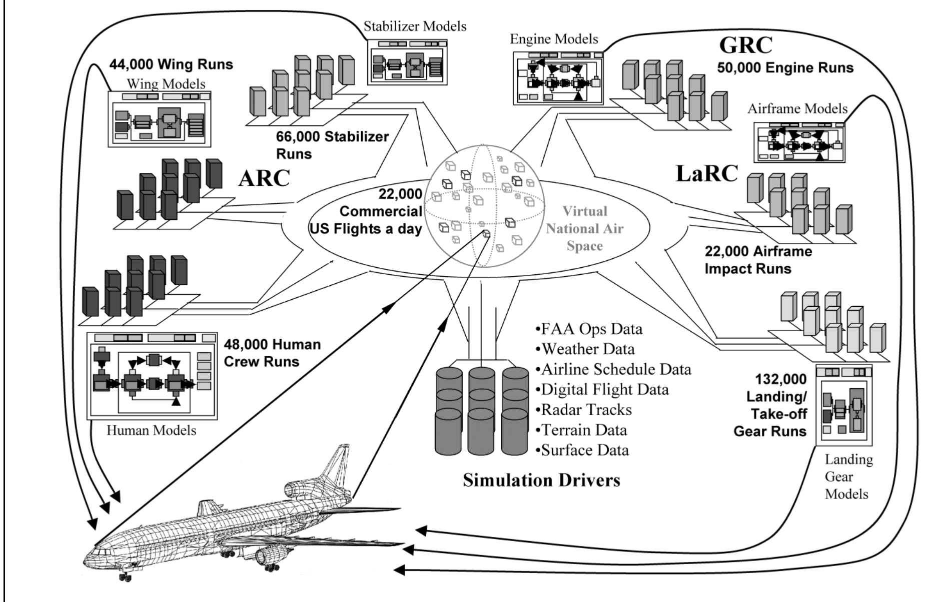

Figure 2. **National Air Space Simulation Environment**

Many aircraft, flight paths, airport operations, and the environment are combined to get a virtual national airspace. (Figure courtesy Yuri Gawdiak, NASA Ames.)

2.2 DOE's Supernova Cosmology Program

Over the past several years, astronomers and astrophysicists have been conducting in-depth sky searches with the goal of identifying certain reference types of supernovae in their earliest evolutionary stages and then, during the two to four weeks of their most "explosive" activity, measuring their changing magnitude and spectra. These "standard candles," as they are called by the astronomers, are supernova that can be used to directly measure various cosmological properties. (See [4] and [5].) These early experiments have demonstrated that the expansion of the universe is accelerating, apparently driven by an unknown new force that overwhelms the force of gravity, contrary to existing models where gravity would cause the universe expansion to slow. The discovery of this new force – now called dark energy – is a stunning discovery and was named the "breakthrough of the year" by Science Magazine in Dec.1998.

These experiments have been daunting tasks in terms of both the number and volume of observations required. The early successes have driven the expansion of these searches in terms of both sky area and apparent magnitude observed. The search program currently under development at LBNL, the Supernova Factor (http://snfactory.lbl.gov), is an earth-based observation program utilizing observational instruments at Haleakala and Mauna Kea, Hawaii and Mt. Polomar, California. When fully implemented, this search program will also utilize instruments at observatories in Chile and the Canary Islands. This program will also serve as a development testbed for the next generation search program, the space-based Supernova Acceleration Probe (SNAP). The Supernova Acceleration Probe is a satellite-based supernova search program combining an optical field imager, near infrared imager and spectrometer in a single, dedicated spacecraft (see http://snap.lbl.gov).

[a] This scenario does not imply complete, high fidelity simulation of every aircraft in the air space. Most of the simulations are likely to be run as low fidelity simulations that can rapidly be converted to high fidelity simulations, if the need arises.

This new approach to cosmology – only possible because of the availability of large-scale computing and data storage facilities at the DOE NERSC facility and the corresponding NSF supercomputer centers – is called "observational cosmology".

The evolution from proof-of-principle to full scale supernova search has unveiled new operational issues for these research programs that we feel are characteristic of how modern science is evolving under the influence of vastly increased distributed computing and data handling capabilities. The first of these is the shear scale of the computing and data-handling task involved. Raw, uncorrected sky images must be transferred nightly from remote observatories to central computing facilities, NERSC [6] in this case. Here, these images undergo extensive computational calibration and correction to eliminate sky tracking errors as well as instrumentation and atmospheric effects. The resulting images must then be compared to recent baseline sky catalogs in order to eliminate asteroids and man-made satellite tracks. Only then can automated search algorithms look for increases in stellar magnitude that may indicate the onset of supernova activity. Fifty plus Gigabytes in some 500 files need to be shepherded through this process of data transfer, computation and archiving on a daily basis for the 5 to 10 years duration of the search effort. The script and operator based automation used during early sky search programs simply will not scale to the levels of performance and reliability required by these new searches.

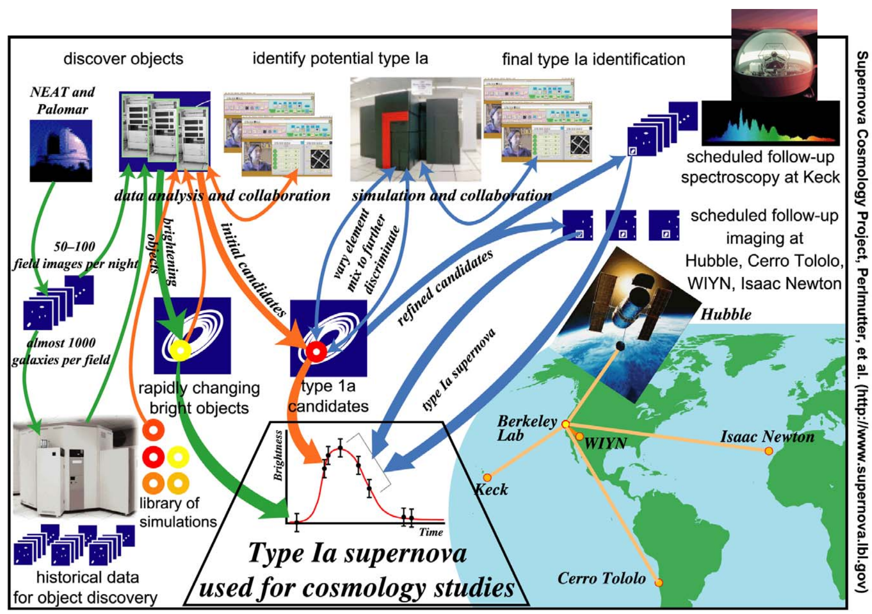

Figure 3. Supernova Cosmology Requires Close Integration of Observation and Data Analysis with Supercomputing

This figure illustrates, left to right, the acquisition and successive refinement of data leading to the observations that provide cosmological information.

Secondly, the amazing experimental results obtained thus far have promoted strong programmatic ties between cosmologists involved in modeling stellar behavior through simulations, and those engaged in direct observation. Simulation teams are now engaged in ambitious efforts to develop new models that provide full 3D simulation of both the hydrodynamic and radiative transfer aspects of supernova that can predict, based on the parameters of the exploding star, the spectra during supernova. Since the development of accurate models requires a detailed comparison with observed supernova data, data from the Supernova Factory is of critical importance in the successful development of these models. Although the initial motivation is the improvement of current computational models through direct and frequent comparison to observations, ultimately the goal is to use closely coupled observation/simulation efforts to filter out supernova candidates that are not the reference types useful for cosmology. As both the number of discovered supernovae and the demands for scarce, shared observational instruments increase, the ability to successfully filter unwanted supernovae out of the observational program becomes increasingly important. This is accomplished by using the initial observation to establish the parameters for the simulations, which, in turn, predict the observed spectra in order to determine the exact type of the supernova.

When this determination of type results in identifying a "standard candle" (type 1a) supernova, this information must be immediately conveyed to one of the large instruments such as Keck, Palomar, or Hubble, in order to observe the spectra throughout the short (weeks long) life of the supernova. (See the right most process illustrated in Figure 3.) This is the information that permits cosmological inference.

The combination of these processes establishes a cycle of coarse observation – simulation – detailed spectra observation that is time constrained by the fact that the useful spectrum observation period is only for a few weeks following discovery.

Largely, this automation is required because of the sheer complexity of the operations involved. Input data and calibration files will need to be staged to disk before analysis programs can begin. Resulting data files will need to be archived, cataloged and published to other collaborating programs. In addition, in the presence software or hardware errors, these activities need to be rescheduled in a carefully controlled sequence to insure their proper completion. Some experiments, notably the Supernova Factory, will require that these operations occur, with a minimum of human intervention, around the clock, whenever observational data from earth and space instruments has been transferred to HPSS at NERSC. This workflow must be managed autonomously and reliably in order to meet the needs of the science.

Security and authorization also acquire significant importance when developing mechanisms that allow collaborators throughout the world to monitor and control daily analysis and archiving efforts. Success will depend on collaborating scientists being able to manage data processing and storage and to integrate advanced supernova simulation into the real-time control of the experiments. The ability to perform real-time control will allow collaborating scientists in one part of the world to look at results and change viewing plans in another part, thus taking advantage of the different time zones across the collaboration – an important aspect when observation can only be done for a few hours each night. This sort of access to the supercomputers and instruments must be able to be done securely or it will not happen.

2.3 Application Characteristics

From these examples, we can enumerate a set of high-level characteristics of the applications.

From the virtual sir space application, we can identify the following.

- system simulations are built up by coupling legacy code components
- computing capacity and simulation expertise will come from many different organizations
- simulation components must be coordinated on many different computing systems
- aircraft simulations must be coupled to the independent environmental and operations data sources that originate from hundreds of different locations
- confidentiality of data and data access policy enforcement is required both for physical security of the aircraft and for protecting airline proprietary flight operations data
- security and access control for the underlying computing and data archive systems must prevent service disruption

Characteristics of the supernova cosmology project include:

- the cycle of coarse observation – simulation – fine observation represents a complex workflow that involves human interaction
- the workflow process involves time constrained use of supercomputers
- the final observations require interaction with potentially on-line instruments
- the scientific teams are loose collaboration of researchers that are distributed world-wide
- sufficient security and access control are required to prevent disruption and un-authorized access to the computing, data, and instrument systems

3 Application Needs and the Role of Grids

The application characteristics from the examples above imply a collection of capabilities that must be addressed by the computer scientists who build the distributed systems that combine the computational tool with data and instruments..

In addition to the environment and services needed to support these applications, our experience in working with the design engineer / analyst who must use the system to accomplish a specific task suggests many other characteristics and requirements as well.. (E.g., see [7].)

The requirements of these several application areas lead to a characterization of the desired Grid functionality. This functionality may be represented as a hierarchically structured set of services and capabilities that are described below, and whose interrelationship is illustrated in Figure 4. Some of the key issues that Grids address now, or will in the near future, include:

- techniques for coupling heterogeneous computer codes, resources, and data sources in ways so that they can work on integrated/coupled problems in order to provide whole system simulations ("multi-disciplinary simulation/optimization")
- comprehensive network monitoring to locate, analyze, and correct bandwidth bottlenecks

- data replica catalogues to provide global views of cached data
- methodology and implementation for incorporating, using, and managing resources in the overall environment that are scalable to thousands of resources
- security and management of access rights for the collaboration data and information

Scientific and Engineering applications involving distributed teams and distributed resources have lead to specific requirements for workflow and collaboration frameworks, including:

- describing and managing multi-step, asynchronous component workflows, including managing fault detection and recovery
- access to data and metadata publication and subscription mechanisms
- event mechanisms - e.g. notification of when data or simulation results come into existence anywhere in the space of resources of interest
- user interfaces to each of the above

4 A Model for Computing and Data Grids

Grid environments that provide the services noted above are structured into a number of “levels”. (See Figure 4.) There are services that provide the user interfaces and application regime workflow management, tools and services supporting the development of application programs, the basic Grid services that provide uniformity and access to resources, and there are the resources themselves. Ancillary services such as security and system management are required at all levels.

An understanding is emerging a to what are the Grid services needed to support large-sacle science and engineering activities. Many of the basic Grid serrvices are currently available, and those are indicated in Figure 4. However, by no means are all of the required services currently available in Grids. Many of the more application oriented services like workflow and collaboration services are still under development.

4.1 Problem Solving Environments: Knowledge Based Queries, User Interfaces and Workflow Management

The Knowledge Grid [a]

It is clear that for the Grid to realize the maximum impact on science and engineering that there must be mechansims for discipline problem solvers to be able to express a problem in terms of the knowledge framework of their discipline, and then have that problem translated to the computational and data analysis operations of the underlying problem solving system. There have been various discipline specific efforts to do this sort of thing, but not much general infrastructure has been developed. The approach of Cannataro, et al [8] suggests one way to approach at least the representation and manipulation of the knolwedge base that could trtanslate moderately abstract queries in to sets of computations and data analysis that resolve the query.

The User Interface: Integration with the Desktop

A number of services directly support using the Grid by engineers or scientists. These include the toolkits for construction of application frameworks / problem solving environments (PSEs) that integrate Grid services and applications into the “desktop” environment. Services available in the user interface should include, for example, the graphical components (“widgets” / applets) for building application user interfaces; methods for control of the computer mediated; distributed human collaboration that support interface sharing and management; the tools that access the resource discovery and brokering services; the tools that provide generalized workflow management services such as resource scheduling and managing high throughput jobs, etc.

All of these services should be available through Web / desktop interfaces in order to produce a highly usable environment. In this environment, problem-solving protocols may be formulated, controlled, modified, and integrated with other aspects of the work environment, and shared securely with collaborators. This sharing should be able to snapshot the current state of the PSE and pass this snapshot, or a functional replica of it, to a collaborator so that the same view of the application can be viewed and potentially manipulated by the collaborators. It should also be possible to use the PSE mechanism to insert “probes” into the workflow in order to monitor and diagnosis the functioning of the application system.

[a] I am indebted to Mario Cannataro, Domenico Talia, and Paolo Trunfio of the Istituto per la Sistemistica e l'Informatica (ISI) Consiglio Nazionale delle Ricerche (CNR), Italy, for this term.

Workflow Management

Reliable operation of large and complex data analysis and simulation tasks requires methods for their description and control. A workflow management system must provide for a rich and flexible description of the analysis processes and their inter-relationships, and also provide mechanisms for fault detection and recovery strategies in widely distributed systems.

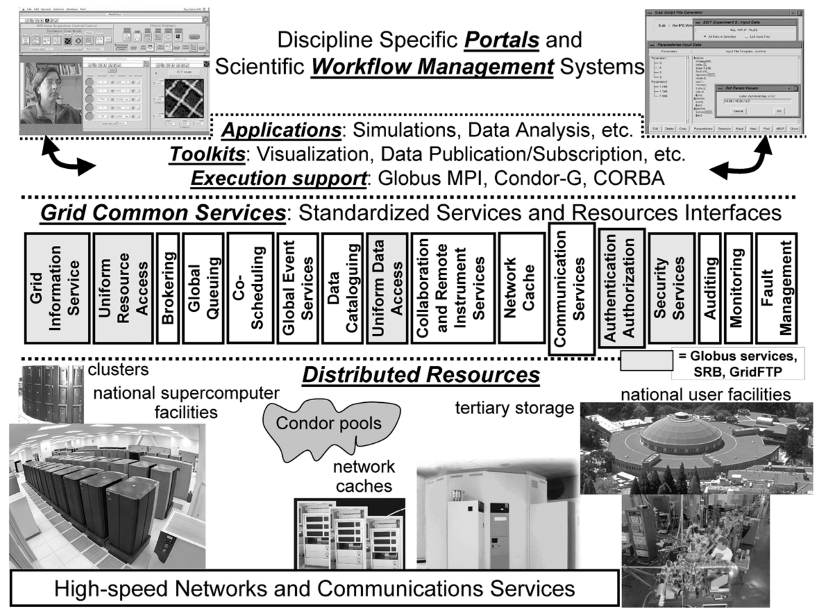

Figure 4. A Grid Architecture

Within the problem solving environments / frameworks must be mechanisms for describing the process of science: The protocols for experiments and hypothesis testing – the definition and management of the interplay of the data generation, data analysis, comparison with simulation, feedback to experiment, etc. Workflow management systems (e.g. as contained in the “framework” in the figure) will carry out the human defined protocols for, e.g., multi-disciplinary simulations and data analysis; global data cataloguing and replica management systems to manage the data for these scenarios, and global event services to manage the dynamic aspects of work protocols, will be essential adjuncts to the workflow engines. That is, these are the services needed directly by scientific and engineering problem solvers.

Collaboration tools

Toolkits supporting the construction of PSEs must also provide the mechanisms for integrating computer mediated, distributed human collaboration into desktop problem solving environments. E.g. interface sharing, graphical user interface components that map to applications and Grid services, access control, a representation of the human work process that maps onto the workflow management mechanism, etc.

Collaboration tools must support the loosely bound collaborations of the scientific community (see [9]). Ideally these would include shared electronic notebooks, tele-meeting tools, tele-presence tools for laboratories and experiment sites, shared authoring tools, shared data publication tools. These tools will be built on Grid services such as secure group communication (“reliable multicast,” e.g., see [10]), which is the basic service for managing distributed, interacting, group services and the Grid Information Service for managing virtual organizations.

4.2 Programming Services

Tools and techniques are needed for building applications and applications systems that are built up from federated components that run in Grid environments. These techniques need to cover a wide spectrum of programming paradigms, and must operate in multi-platform, heterogeneous computing environments. For example, Grid enabled MPI [11] to support the IPC typical of numerical computations; Java and Python bindings to Grid services; CORBA integrated with Grid services to support building CORBA frameworks (like NPSS mentioned above), that federate application components; Condor-G [12] for managing large numbers of related jobs such as parameter studies and data analysis; Java/RMI and DCOM to obtain access to various commercial services; are all application oriented middleware systems that will have to interoperate with the Grid services in order to gain access to the resources managed by the Grid.

4.3 Grid Common Services

“Grid Common Services” refers to the basic services that provide uniform and location independent access and management of distributed resources. Much of the operational effort to run Grids is involved in maintaining these services.

Many Grids (including NASA's IPG [13], DOE's Science Grid [14], and ASCI Grid [15], and the Grids of the NSF Supercomputer centers[16]) currently use Globus [17] to provide the basic services that characterize and locate resources, initiate and monitor jobs, provide secure authentication of users, provide uniform access to data, etc.

Grid Information Service

The Grid Information Service has the task of representing, and/or providing access to, virtually all aspects of the configuration and state of the Grid: resource characteristics, virtual organization scoping, presistent data catalogue locations, etc. It's basic function is to be able to respond to queries about the avaiability of resources with certain characteristics and performance state, e.g. the computing systems architecture needed for a particular code. In large-scale environments it may also have the function of providing a rooted namespace into which can be inserted links to other directory services, such as definitions of virtual organizations, locations of data and data replica catalogues for discipline specific data, etc. The GIS must also provide for installing new objects/services into the Grid and must make these new objects known. In this role, the GIS also provides the mechanisms for defining the relationships among Grids, providing a framework for federating Grids, for administrative scoping, etc.

This service – currently provided by the Globus, Grid Information Service [18] – maintains detailed characteristics and state information about all resources, and will also need to maintain, or provide pointers to services that provide, dynamic performance information, information about current process state, user identities, allocations and accounting information.

Execution Management

Several services are critical to managing the execution of application codes in the Grid. The first is resource discovery and brokering in order to build the (usually distributed and trensient) platform – the ensemble of computing, data stroage, etc., systems needed to support the application. By discovery we mean the ability to find the set of objects (e.g. databases, CPUs, functional servers) with a given set of properties that are needed by a distribtued application system. Once the potential resources are identified by such queries to the GIS, the selection of the resourcs to actuallt be used is based on constraints such as allocation and scheduling is a brokering function that will be built on the GIS services. The second is execution queue management, which relates to global views of CPU queues and their user-level management tools (Condor-G is an example an execution management tool). The third category is distributed application management. The last category includes tools for generalized fault management, for monitoring, and, e.g., supplying information to knowledge based recovery systems in the workflow management system.

Runtime

Runtime services include, e.g., checkpoint/restart mechanisms, access control, a global file system, and Grid communication libraries (such as a network-aware MPI) that support security, group communication (reliable multicast) and remote I/O.

Uniform naming and location transparent access must be provided for resources such as data objects, computations, instruments and networks. Transparent access requires uniform I/O mechanisms (e.g. read, write, seek) for all access protocols (e.g. http, ftp, nfs, Globus Access to Secondary Storage, etc.) and richer access and I/O mechanisms (e.g. "application level paging") that are present in existing systems. Currently GridFTP [19] and MCAT/SRB [20] are providing some of these services.

High-speed, wide area, access to tertiary storage systems will always be critical for the science and engineering applications that we are addressing. High-performance applications require high-speed access to data files, and the Grid services must be able to stage, cache, and automatically manage the location of local, remote and cached copies of files. We are also going to need the ability to dynamically manage large, distributed "user-level" caches and "windows" on off-line data. Support for object-oriented data management systems will also be needed. Several of these services will become available over the next year or so from the GriPhyN (Grid Physics Network) project [21].

Services supporting collaboration and remote instrument control, such as secure, reliable group communication are needed. In addition, application monitoring and application characterization, prediction, and analysis, will be important for both users and the managers of the Grid. The NetLogger toolkit [22] and the Network Weather

Service [23] are both being integrated as Grid services, and the GridForum, Grid Performance Working Group [24] is addressing this issue in a general way.

Monitoring services will need precision time event tagging for distributed, multi-component performance analysis. Generalized auditing of data file history and control flow tracking in distributed, multi-process simulations will be needed for integrity, change tracking, fault recovery, and security. General Grid event services are being addressed in the Grid Forum, Grid Computing Environments working group (see the GCE working group pages at www.gridforum.org [25]).

4.4 Resource Management for Co-Scheduling and Reservation

One of the most challenging and well known Grid problems is that of scheduling scarce resources such as supercomputers and large instruments. In many, if not most, cases the problem is really one of co-scheduling multiple resources. Any solution to this problem must have the agility to support transient applications based on systems that are built on-demand for limited periods of time, and in the case of Grid applications that analyze data from scientific and engineering experiment, the Grid resources are likely to have to be available on the sechedule of the instruments. In other words, not only will resources have to be co-secheduled, but they must be scheduled for a paticular time and date. CPU advance reservation scheduling and network bandwidth advance reservation are critical components to the co-scheduling services. In addition, tape marshaling in tertiary storage systems to support temporal reservations of tertiary storage system off line data and/or capacity is likely to be essential, and some of this is provided, e.g., by HRM [26]. The basic functionality for co-scheduling and/or resource reservation must almost always be provided by the individual resource managers, however Grid services serve to coordinate and provide uniform access to these resource specific services.

CPU advance scheduling services are currently provided, e.g., by PBSPro (http://www.pbspro.com/), and is a topic in the Grid Forum's Scheduling Working Group (see [25]).

4.5 Access Control and Security

The first requirement for establishing a workable authentication and security model for the Grid is to provide a single-sign-on authentication for all Grid resources based on cryptographic credentials that are maintained in the users desktop / PSE environment(s) or on one's person. This is provided by X.509 identity certificates or Kerberos credentials, together with the Globus proxies and the services that use them. See [27]. In addition, end-to-end encrypted communication channels are needed in for many applications in order to ensure data integrity and confidentiality.

The second requirement is an authorization and access control model that provides for management of stakeholder rights (use-conditions) and trusted third parties to attest to corresponding user attributes. A policy-based access control mechanism that is based on use-conditions and user attributes is also a requirement. Several approaches are being investigated for providing these capabilities (see, e.g., [28]) and work is being done on integrating these with Grids.

4.6 Services for Operability: Operations and System Administration

Implementing a persistent, managed Grid requires tools for deploying and managing the system software. In addition, tools for diagnostic analysis and distributed performance monitoring are required, as are accounting and auditing tools. Operational documentation and procedures are essential to managing the Grid as a robust production service.

To operate the Grid as a reliable, production environment is a challenging problem. Some of the identified issues include management tools for the Grid Information Service; diagnostic tools so operations/systems staff can investigate remote problems, and; tools and common interfaces for system and user administration, accounting, auditing and job tracking. Verification suites, benchmarks, regression analysis tools for performance, reliability, and system sensitivity testing are essential parts of standard maintenance.

Tools and documentation for operating production Grids are being developled at NCSA [29], and in the IPG [13] and DOE Science Grid [14] projects.

4.7 The Architecture of Grids: How do all these services fit together?

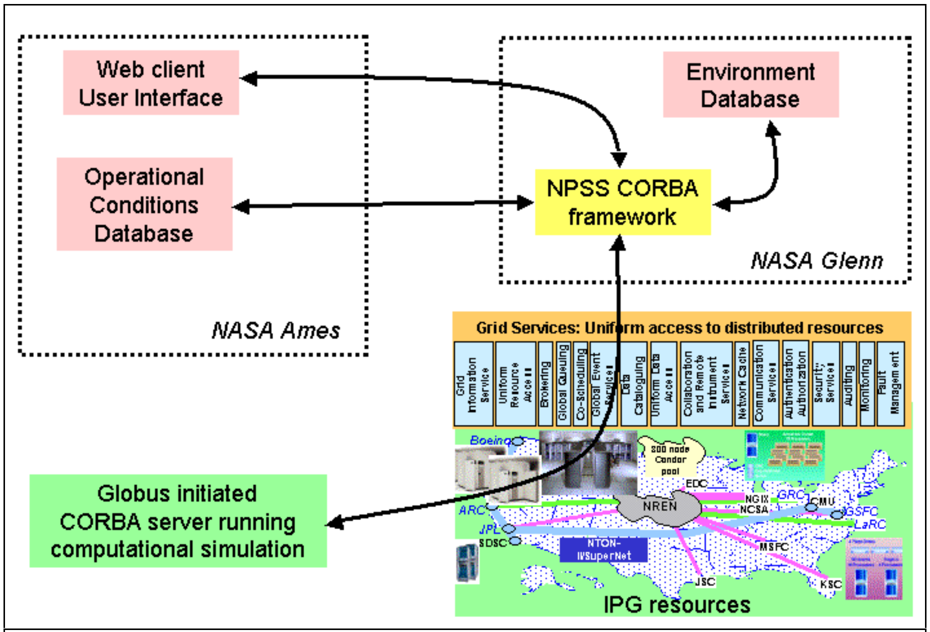

Figure 5. **NPSS / CORBA use of Grids**

We conceptualize the Grid as a layered set of services, as illustrated in Figure 4, that manage the underlying resources, and middleware that supports different styles of usage (e.g. different programming paradigms and access methods).

However, the implementation is that of a continuum of somewhat hierarchically related, independent and interdependent services, each of which performs a specific function, and may rely on other Grid services to accomplish its function.

Further, the “layered” model should not obscure the fact that these “layers” are not just APIs and their underlying protocols, but usually a collection of functions and management systems that work in concert to provide the “service” at a given “layer.” The layering cannot be, and is not, rigid. “Drill down” (e.g. code written for specific system architectures and capabilities) must be easily managed by the Grid services.

Many of these services, and indeed the Grid architecture itself, are the subject of work in the Global Grid Forum, Grid Protocol Architecture Working Group [30]. Also see [31].

5 Current State of Grids

There are several Grids that are at, or close, to production status. Here we describe NASA’s Information Power Grid (www.ipg.nasa.gov).The DOE Science Grid (www.doesciencegrid.org) and the UK eScience Grid (www.nesc.ac.uk/intro) are also working on production Grid environments.

IPG has, over the past three years, deployed a prototype production Grid. By production, we mean that the services and resources are persistent, there are operational groups responsible for those services and resources, and there is documentation and user support.

In the process of building this Grid environment a great deal is being learned about integrating Grids into production supercomputing environments, and some of the issues and lessons learned are documented in [32].

6 Application Use of Grids

In this section, we present two examples of current applications use of Grids. These examples illustrate some Grid successes.

- Standardized access to multi-institutional resources
- A common security approach and infrastructure
- Persistent Grid services (Globus job management and scheduling) that are used to run application frameworks on an as-needed basis
  - CORBA
  - CONDOR job manager (“Glide-in”)
  - Agent systems / servers (data mining example)

The ability to use the Grid to instantiate application frameworks allows users great flexibility in building their applications in the framework of their choice. They do not have to rely on that framework being provided as a persistent service on all of the computing systems where they need to run – they can instantiate their own environment using persistent Grid services.

6.1 Aviation Safety Distributed Simulation

The NPSS system is a CORBA framework [3] in which the model components and the data are manipulated to solve various engine scenarios. The framework is indicated schematically in Figure 5. The framework data paths and use of Globus for instantiating and managing the CORBA environment on supercomputers are indicated in the figure.

6.2 Data Mining

The University of Alabama in Huntsville has developed a data mining system called ADaM (Algorithm Development and Mining). The current design consists of a mining engine and a daemon-controlled database. The database contains information about the data to be mined including its type and its location. To mine for data, the user provides the mining engine with a mining plan that consists of the sequential list of mining operations that are to be performed along with any parameters that may be required for each mining operation. The mining engine consults the database in order to find out where the data to be mined is stored and then applies the mining plan to the set of data that has been identified to the database. Each mining operation is represented as a shared-library file, one file per operation.

The IPG version of ADaM is structured so that the database and its associated daemon resides on a processor distinct from where the mining engine operates. The data is managed at multiple sites by SRB/MCAT and GridFTP.

See [33].

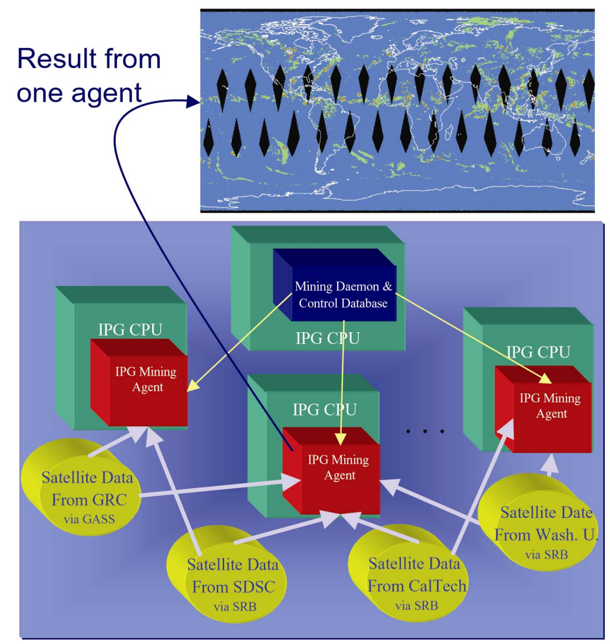

Figure 6. Globus Managed, Agent Based Data Mining Framework

512 node SGI Origin at NASA Ames uses IPG uniform interface data access tools to simultaneously mine hydrology data from four sites.

7 Future Directions

Much of the current use of Grids falls into two categories. One is to instantiate a framework on the remote machine, and then proceed within that framework. The NPSS and data mining applications described above are examples of this. The second is to manage large numbers of small jobs (parameter studies and data analysis), and the ILAB systems described above is an example of this.

While these are early successes in using Grids, they are only a first step to the level of Grid technology and deployment that we need to have a substantial impact on large-scale science and engineering.

Another area that will be critical for Grids to facilitate applications with very large data handling problems such as the high-energy physics experiments described above. There are several "Data Grid" projects whose whole purpose is to address Grid technologies for this massive data handling, and the reader is referred to [21] which provides a very nice overview of the issues and approach.

8 Acknowledgements

IPG is funded primarily by NASA's Aero-Space Enterprise, Information Technology (IT) program (http://www.nas.nasa.gov/IT/overview.html). DOE's Science Grid is funded by the U.S. Dept. of Energy, Office of

Science, Office of Advanced Scientific Computing Research, Mathematical, Information, and Computational Sciences Division (http://www.sc.doe.gov/production/octr/mics) under contract DE-AC03-76SF00098 with the University of California. This paper is LBNL-49741.

9 Notes and references

[1] *The Grid: Blueprint for a New Computing Infrastructure*, I. Foster and C. Kesselman, eds. 1998, Morgan Kaufmann. http://www.mkp.com/books_catalog/1-55860-475-8.asp

[2] "Numerical Propulsion System Simulation," NPSS. http://hpcc.lerc.nasa.gov/npssintro.shtml

[3] "A CORBA-based Development Environment for Wrapping and Coupling Legacy Codes," G. Follen, C. Kim, I. Lopez, J. Sang and S. Townsend. In *Tenth IEEE International Symposium on High Performance Distributed Computing*. 2001. San Francisco.

[4] "Supernova Cosmology Project," S. Perlmutter and e. al. 2001. http://www.supernova.lbl.gov

[5] The "Cosmology Tutorial" is a good introduction to cosmology in general, and also specifically discusses supernova cosmology (at http://www.astro.ucla.edu/~wright/sne_cosmology.html) "Cosmology Tutorial," N. Wright. 2001, UCLA. http://www.astro.ucla.edu/~wright/cosmolog.htm

[6] NERSC is one of the largest unclassified scientific supercomputer centers in the US. It's mission is to accelerate the pace of scientific discovery in the DOE Office of Science community by providing high-performance computing, information, and communications services. NERSC is the principal provider of high performance computing services to Office of Science programs -- Magnetic Fusion Energy, High Energy and Nuclear Physics, Basic Energy Sciences, Biological and Environmental Research, and Advanced Scientific Computing Research. "National Energy Research Scientific Computing Center," NERSC. 2001. www.nersc.gov

[7] "Information Power Grid Implementation Plan: Research, Development, and Testbeds for High Performance, Widely Distributed, Collaborative, Computing and Information Systems Supporting Science and Engineering," W. Johnston, D. Gannon and B. Nitzberg. 2001: NASA Ames. http://www.ipg.nasa.gov/engineering/requirements/requirements_index.htm

[8] "The Knowledge Grid: Towards an Architecture for Knowledge Discovery on the Grid," M. Cannataro, D. Talia and P. Trunfio. In *FIRST EUROGLOBUS WORKSHOP*. 2001. Robinson Club Apulia Village, Marina di Ugento, Lecce. http://www.euroglobus.unile.it/

[9] The PCCE is a web-based persistent space that supports continuous and ad hoc collaboration. It targets daily tasks such as document sharing and provides a base connectivity for communication. Documents can be published into the web environment (by users or applications) and viewed in a browser via WebDAV capabilities. "Pervasive Collaborative Computing Environment," D. Agarwal. http://www-itg.lbl.gov/Collaboratories/pcce.html

[10] "Reliable and Secure Group Communication," D. Agarwal, K. Berket and O. Chevassut. 2001. http://www-itg.lbl.gov/CIF/GroupComm

[11] "A Grid-Enabled MPI: Message Passing in Heterogeneous Distributed Computing Systems," I. Foster and N. Karonis. In *SC98*. 1998. http://www.globus.org/research/papers.html#mpich98

[12] "Condor-G," J. Frey, T. Tannenbaum, M. Livny, I. Foster and S. Tuecke. In *Proceedings of the Tenth International Symposium on High Performance Distributed Computing (HPDC-10)*. 2001: IEEE Press. http://www.globus.org/research/papers.html#Condor-G-HPDC

[13] "NASA's Information Power Grid," IPG. http://www.ipg.nasa.gov

[14] The DOE Science Grid's major objective is to provide the advanced distributed computing infrastructure based on Grid middleware and tools to enable the degree of scalability in scientific computing necessary for DOE to accomplish its missions in science. "DOE Science Grid," Science_Grid. 2001. http://www-itg.lbl.gov/Grid/

[15] "Constructing the ASCI Grid," J. Beiriger, W. Johnson, H. Bivens, S. Humphreys and R. Rhea. In *Proc. 9th IEEE Symposium on High Performance Distributed Computing*. 2000: IEEE Press.

[16] The NSF PACIs are the Alliance/NCSA (http://www.ncsa.uiuc.edu/) and NPACI/SDSC (http://www.npaci.edu/) "PACI," PACI.

[17] "Globus: A Metacomputing Infrastructure Toolkit", I. Foster and C. Kesselman. Int'l J. Supercomputing Applications, 1997. http://www.globus.org/research/papers.html

[18] "Grid Information Services for Distributed Resource Sharing," K. Czajkowski, S. Fitzgerald, I. Foster and C. Kesselman. In *High Performance Distributed Computing* **-** *10*. 2001. San Francisco, CA. http://www.globus.org/research/papers.html#MDS-HPDC

[19] "GridFTP: Universal Data Transfer for the Grid", Globus Project. 2001. http://www.globus.org/datagrid/

[20] "Massive Data Analysis Systems," R. Moore and et al., San Diego Supercomputer Center. http://www.sdsc.edu/MDAS

[21] The GriPhyN (Grid Physics Network) collaboration is a team of experimental physicists and information technology (IT) researchers who plan to implement the first Petabyte-scale computational environments for data intensive science in the 21st century. Driving the project are unprecedented requirements for geographically dispersed extraction of complex scientific information from very large collections of measured data. To meet these requirements, which arise initially from the four physics experiments involved in this project but will also be fundamental to science and commerce in the 21st century, GriPhyN will deploy computational environments called Petascale Virtual Data Grids (PVDGs) that meet the data-intensive computational needs of a diverse community of thousands of scientists spread across the globe. "GriPhyN (Grid Physics Network) project proposal," GriPhyN. 2000. http://www.griphyn.org/info/itr2000.html

[22] "NetLogger: A Toolkit for Distributed System Performance Analysis," D. Gunter, B. Tierney, B. Crowley, M. Holding and J. Lee. In *IEEE Mascots 2000: Eighth International Symposium on Modeling, Analysis and Simulation of Computer and Telecommunication Systems*. 2000. http://www-didc.lbl.gov/papers/NetLogger.Mascots.paper.ieee.pdf

[23] "Forecasting Network Performance to Support Dynamic Scheduling Using the Network Weather Service," R. Wolski, in *Proc. 6th IEEE Symp. on High Performance Distributed Computing*. 1997: Portland, Oregon.

[24] The Grid Performance working group is focused on defining standards and best practices for the gathering, representation, storage, distribution, and query of performance information about Grid resources and applications. R. Aydt, D. DGunter and W. Smith. 2001. http://www-didc.lbl.gov/GridPerf/

[25] The Global Grid Forum (www.Gridforum.org) is an informal consortium of institutions and individuals working on wide area computing and computational Grids: the technologies that underlie such activities as the NCSA Alliance's National Technology Grid, NPACI's Metasystems efforts, NASA's Information Power Grid, DOE ASCI's DISCOM program, and other activities worldwide. Grid_Forum.

[26] "Storage Management for High Energy Physics Applications," A. Shoshani, L. M. Bernardo, H. Nordberg, D. Rotem and A. Sim. In *Computing in High Energy Physics 1998 (CHEP 98)*. 1998. http://gizmo.lbl.gov/~arie/download.papers.html

[27] "A Security Architecture for Computational Grids," I. Foster, C. Kesselman, G. Tsudik and S. Tuecke, in *ACM Conference on Computers and Security*. 1998. p. 83-91.

[28] "Certificate-based Access Control for Widely Distributed Resources," M. Thompson, W. Johnston, S. Mudumbai, G. Hoo, K. Jackson and A. Essiari. In *Eighth Usenix Security Symposium*. 1999. http://www-itg.lbl.gov/Akenti/papers.html

[29] "Alliance Computational Environments & Security," NCSA. 2001. http://www.ncsa.uiuc.edu/TechFocus/Projects/

[30] The role of the Grid Protocol Architecture Working Group is to provide a conceptual framework for discussing the interrelationships, completeness, and minimality of the protocol approach to Grid services that is coming out of GF. "Global Grid Forum, Grid Protocol Architecture Working Group," W. Johnston. 2001. http://www-itg.lbl.gov/GPA

[31] "The Anatomy of the Grid: Enabling Scalable Virtual Organizations", I. Foster, C. Kesselman and S. Tuecke. Intl. J. Supercomputer Applications, 2001. **(to appear)**. http://www.globus.org/research/papers/anatomy.pdf

[32] "NASA's Information Power Grid: Production Grid Experience with Distributed Computing and Data Management," W. Johnston. In *Second Global Grid Forum Workshop (GGF2)*. 2001. Washington, D.C. http://www.itg.lbl.gov/~johnston/Grids

[33] "Data Mining on NASA's Information Power Grid," T. Hinke and J. Novonty. In *Ninth IEEE International Symposium on High Performance Distributed Computing*. 2000.